

Multiple-station data processing

Anna Platz^{1,2} (aplatz@gfz-potsdam.de), Ute Weckmann^{1,2}

¹ Helmholtz Centre Potsdam - German Research Centre for Geosciences GFZ, Potsdam, Germany

² University of Potsdam, Institute of Geosciences, Potsdam, Germany

SUMMARY

Magnetotellurics (MT) relies on natural electromagnetic field variations to investigate the electrical conductivity structure of the subsurface. Modern MT data are often multivariate due to simultaneous recordings of multiple-channel time series of two (horizontal) electric and three magnetic field components at multiple stations. Single site and remote reference processing only use a small portion of data to estimate the impedance tensor. The multiple-station approach, initially presented by Egbert (1997), uses all data information to improve the signal-to-noise ratios, which results in better estimations of the transfer functions. This is particularly important in industrialized regions, where the influence of man-made noise signals often exceeds the natural EM fields and hampers the estimation of MT impedance tensors. We have included the multiple-station data approach in our processing scheme EMERALD and tested it with different data sets. A non-robust calculation of the impedance tensor based on the multiple-station approach shows already slightly improved results compared to robust single site or even remote reference estimators. However, in case of high level man-made noise advances of the multiple-station algorithm are not observed. Tests with existing robust routines within EM, which are based on bivariate assumptions, do not reveal a significant improvement.

Keywords: Magnetotellurics, multiple-station data processing

INTRODUCTION

The magnetotelluric method (MT) utilizes natural electromagnetic (EM) field variations, which penetrate into the subsurface and induce secondary fields depending on the electrical conductivity structure of the Earth. In industrialized regions these natural signals are (severely) disturbed by man-made noise sources, which often exceed the natural EM fields and hamper the estimation of MT impedance tensor. To obtain high quality MT results advanced processing approaches have to be applied. The multiple-station approach, initially presented by Egbert (1997), uses all data information to improve the signal-to-noise ratios and provides indications for the existence of coherent noise. Furthermore the approach works in the frequency domain with Fourier coefficients derived from a series of short overlapping time segments and uses a robust MEV (multivariate errors-in-variables) estimation scheme. This multiple-station approach has been included in our processing routines; however, having an interface based on auto and cross spectra averaged to specific centre frequencies, it has to be modified and tested with parts of our robust algorithm in EMERALD, being different from Egbert's robust algorithm.

Six stations, measured in Namibia in 2011, are used to test the new data processing approach within

EMERALD. We assigned three stations to a network, marked with red and blue stars in Figure 1. All of these stations have high data quality and can be regarded as a reference data set for testing the approach. We will show results of the common single site processing compared with multiple-station processing results and discuss the influence of the applied robust statistics scheme within EMERALD.

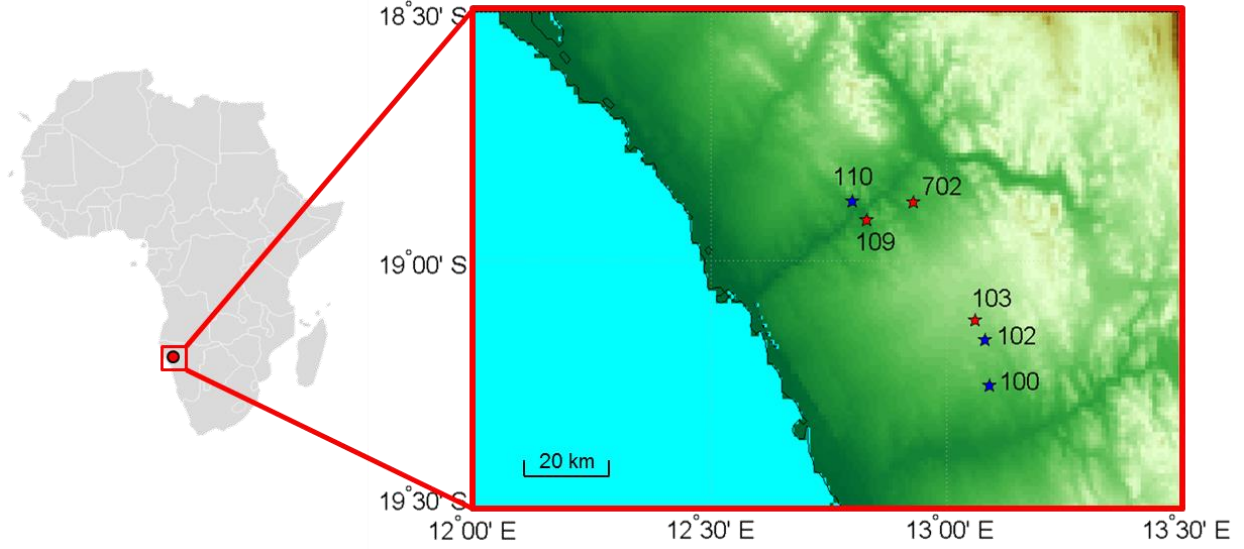


Figure 1: Locations of six exemplary stations from measurements in Namibia: The first network is composed of stations 103, 109 and 702 (red stars) and is characterized by almost noise-free MT data. The second network consists of stations 100, 102 and 110 (blue stars) and contains more EM noise.

NORMAL DATA PROCESSING METHODS

In frequency domain, a linear relationship between the horizontal electric and horizontal magnetic components exists given by the MT impedance tensor \underline{Z}

$$\begin{pmatrix} E_x \\ E_y \end{pmatrix} = \begin{pmatrix} Z_{xx} & Z_{xy} \\ Z_{yx} & Z_{yy} \end{pmatrix} \cdot \begin{pmatrix} B_x \\ B_y \end{pmatrix} \quad (1)$$

with \mathbf{E} being the electric field in $mVkm^{-1}$ and the magnetic field \mathbf{B} in nT . The components Z_{ij} ($i, j = x, y$) of the impedance tensor are in units of ms^{-1} .

For presentation purposes and to convert the complex numbered impedance tensor elements into physically more meaningful quantities apparent resistivity $\rho_{a,ij}$ in Ωm and phase values ϕ_{ij} in deg are computed.

$$\rho_{a,ij}(\omega) = 0,2 T |Z_{ij}(\omega)|^2 \quad (2)$$

$$\phi_{ij}(\omega) = \tan^{-1} \left(\frac{Im\{Z_{ij}(\omega)\}}{Re\{Z_{ij}(\omega)\}} \right) \quad (3)$$

The period length T is given in units of s and the angular frequency ω is expressed in units of Hz . The complex impedance tensor contains information on the conductivity structure of the subsurface and is a final result of data processing. A flow chart of the main steps of the applied data processing EMERALD is shown in Figure 2. The original time series are band-pass filtered into narrow frequency bands and divided into short,

subsequent segments of a fixed length of typically 128 samples. These short segments are cosine tapered prior to Fourier transformation. After which they are called ‘events’. These events are corrected for instrument response functions and then averaged at center frequencies, which are equally distributed on a logarithmic scale (Weckmann et al., 2005). For each sub-band and each of the five measured components smoothed auto and cross spectra are calculated. The estimations of impedance tensor are derived by stacking single event spectra from all frequency bands using a robust algorithm described in Ritter et al. (1998). The results are influenced by coherent and incoherent noise similarly. The magnetic channels are assumed to be noise-free and systematic underestimation of noise in these channels leads to a bias of the apparent resistivity curves since noise does not cancel out in auto spectra. For the remote reference (RR) method, auto spectra in the denominators are substituted by cross spectra of the magnetic channels of the local and the reference site. Therefore the existence of a remote station is mandatory, which has to be located far away from the local station to guarantee that EM noise at both stations is incoherent. Finding and maintaining such a reference site during a field campaign is usually expensive and time consuming.

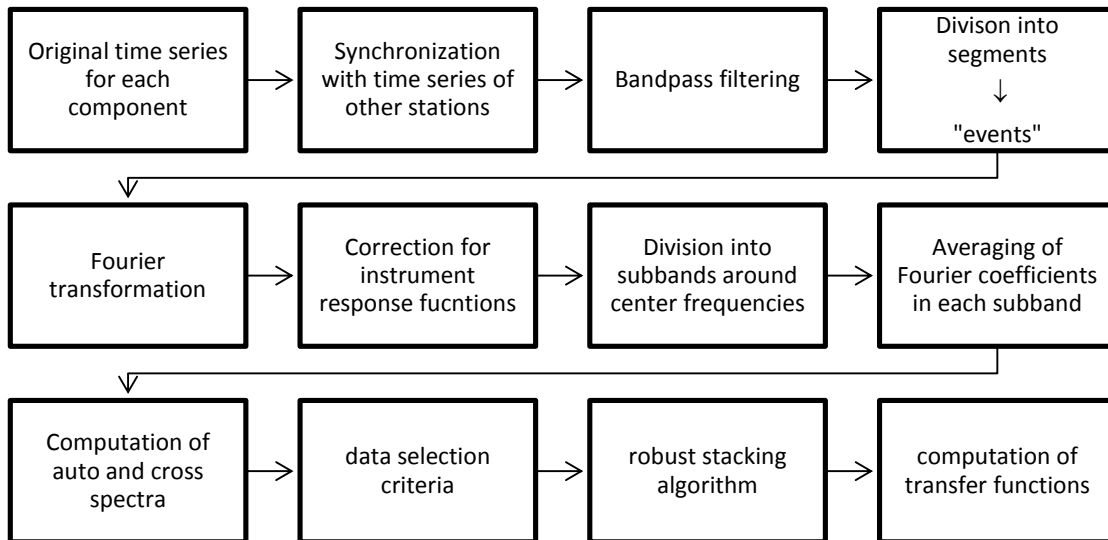


Figure 2: Flow chart of magnetotelluric data processing. All steps displayed in the first row take place in time domain, while the rest already applies to frequency domain.

In Figure 3 results of single site processing for station 109 are shown. The coherency criterion within the robust algorithm is used with a threshold of 0.9. In Figure 3b single site results are overlaid with RR results of station 109 with station 103. The diagonal component Z_{xx} of the impedance tensor shows a typical bias in the period range 1 – 10s, which can be eliminated using the remote reference method (see orange curves in Figure 3b, which overlay the red curves for most frequencies except the frequency range around the dead band between 1s and 10s).

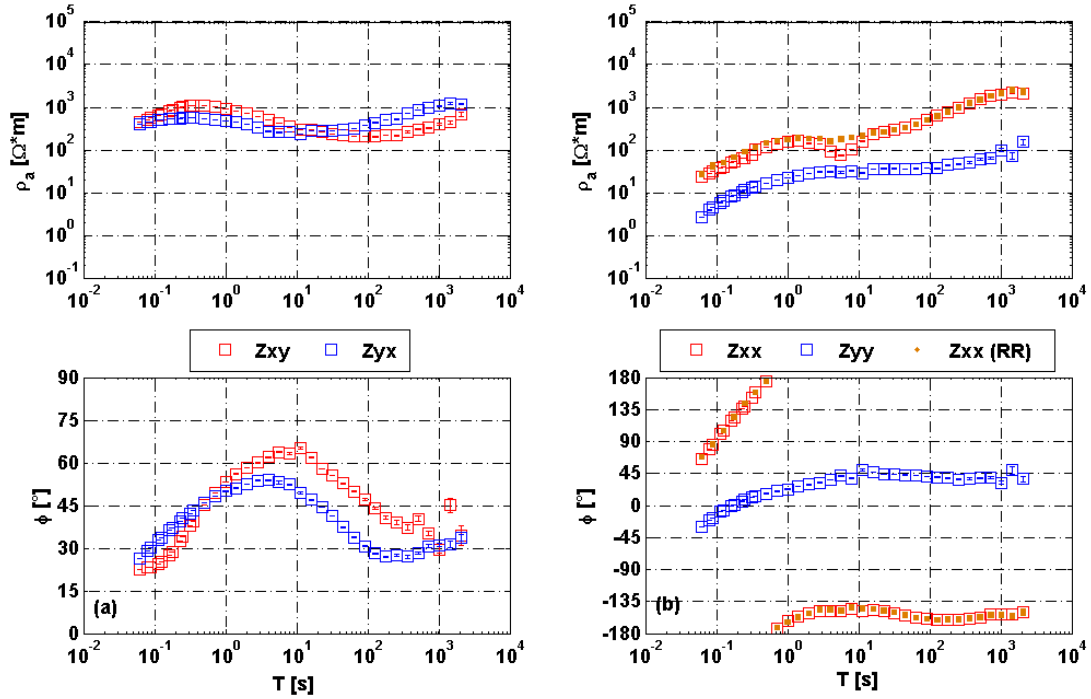


Figure 3: Single site results of station 109. The curves of the off-diagonal components are smooth (a). (b) The apparent resistivity of the Z_{xx} component shows a bias between 1s and 10s, which is eliminated by the remote reference method (see orange curve). Single site as well as remote reference results are obtained by using the robust algorithm within EMERALD with a threshold of 0.9 for the coherency criterion and without phase criterion.

Both methods, single site and remote reference, are based on univariate and bivariate statistics and only use a small portion of the available data. The spectral density matrix (SDM) contains all possible auto and cross spectra, exemplary shown for a single station in (4).

$$\underline{SDM} = \begin{bmatrix}
 \langle B_x B_x^* \rangle & \langle B_x B_y^* \rangle & \langle B_x B_z^* \rangle & \langle B_x E_x^* \rangle & \langle B_x E_y^* \rangle \\
 \langle B_y B_x^* \rangle & \langle B_y B_y^* \rangle & \langle B_y B_z^* \rangle & \langle B_y E_x^* \rangle & \langle B_y E_y^* \rangle \\
 \langle B_z B_x^* \rangle & \langle B_z B_y^* \rangle & \langle B_z B_z^* \rangle & \langle B_z E_x^* \rangle & \langle B_z E_y^* \rangle \\
 \langle E_x B_x^* \rangle & \langle E_x B_y^* \rangle & \langle E_x B_z^* \rangle & \langle E_x E_x^* \rangle & \langle E_x E_y^* \rangle \\
 \langle E_y B_x^* \rangle & \langle E_y B_y^* \rangle & \langle E_y B_z^* \rangle & \langle E_y E_x^* \rangle & \langle E_y E_y^* \rangle
 \end{bmatrix} \quad (4)$$

The SDM is a hermitian matrix, wherefore in our example 10 of the 25 components are dependent coloured in gray. For the estimation of the full impedance tensor only the seven components in blue are used. Most of the calculated spectra therefore are not used for the estimation of the impedance tensor.

Although modern MT data are often multivariate due to simultaneous recordings of multiple-channel time series of two (horizontal) electrical and three magnetic field components at multiple stations, data of each station is processed separately. Unfortunately, in industrialized regions the influence of man-made noise signals often exceeds the natural EM fields and hampers the estimation of MT impedance tensors. This coherent noise can be eliminated neither by remote reference nor by single site processing using robust statistics to obtain a mean impedance tensor. To obtain high quality MT results different processing approaches have to be applied, which utilize available information about signal and noise in all recording

channels. The multiple-station approach, initially presented by Egbert (1997), uses all data information to improve the signal-to-noise ratios by separating different field components and thus hoping to separate and identify EM noise within. Therefore the multiple-station approach works with networks of stations instead of single stations.

MULTIPLE-STATION DATA PROCESSING

The multiple-station approach was presented by Egbert (1997) and works with Fourier coefficients derived from time series. We have included a non-robust multiple-station data processing in our processing routines within EMERALD. The approach is still based on the ideas of Egbert, but we have to adopt it to work with auto and cross spectra. This multiple-station approach is based on a multivariate linear model, which describes the formation of the data vector \vec{X}_i by the electric and magnetic Fourier coefficients h_{ji}, e_{ji} for the i th time segment at the site j of the station network.

$$\vec{X}_i = \begin{pmatrix} h_{1i} \\ e_{1i} \\ \vdots \\ h_{ji} \\ e_{ji} \end{pmatrix} = \begin{pmatrix} \eta_{11} \\ \zeta_{11} \\ \vdots \\ \eta_{j1} \\ \zeta_{j1} \end{pmatrix} \beta_{1i} + \begin{pmatrix} \eta_{12} \\ \zeta_{12} \\ \vdots \\ \eta_{j2} \\ \zeta_{j2} \end{pmatrix} \beta_{2i} + \varepsilon_i = \underline{U} \vec{\beta}_i + \vec{\varepsilon}_i \quad (5)$$

For MT data each station typically consists of two electrical and three magnetic channels, so that in total $K = 5J$ channels exist within a network. The data vector can be sub-divided into three vectors, in which two of them contain the Fourier coefficients η_{ji}, ζ_{ji} corresponding to the two orthogonal plane-wave MT sources. The parameters $\beta_{li}, l = 1, 2$ define the polarization of these source fields and the additional vector $\vec{\varepsilon}_i$ represents all sources of coherent and incoherent noise, which we assume to be statistically independent of our MT signals. The first two vectors can be summarized to the matrix U , which ideally contains the electrical and magnetic fields that would be observed at all sites within a network for idealized quasi-uniform magnetic sources, which are linearly polarized in north-south and east-west direction. All channels of all stations within a network can contain noise, wherefore bias, which is observed in single site processing, is avoided. Coherent noise in the data within a station network can be clearly detected. The impedance tensor for each station can be determined from the elements of U , which correspond to the x and y components of the electrical and magnetic fields observed at the j th station. Estimates of the matrix U can be obtained by the solution of following generalized eigenvalue problem:

$$Su = \lambda \Sigma_N u \quad (6)$$

The matrix S is the spectral density matrix containing in our case all possible auto and cross spectra within a station network and the matrix Σ_N is the covariance matrix of all noise sources. The eigenvectors of the two largest eigenvalues build the matrix U . Unfortunately, in most cases the coherent noise structure within a network is unknown and without a priori knowledge of noise geometry, development of a parametric model for Σ_N is impossible. The simplest model for Σ_N is a diagonal matrix, which assumes that noise is incoherent between stations as well as between channels of a single site.

$$\Sigma_N = \text{diag}(\sigma_1^2 \dots \sigma_k^2) \quad (7)$$

The parameters σ_k^2 represent the variances of incoherent noise and can be calculated from the auto and

cross spectra by using a multiple linear regression to fit data for channel k to the remaining $K - 1$ channels. By definition the incoherent part of the data in each channel is independent from data in the remaining channels, so that the magnitude of the residuals of the multiple linear regression provides an estimation of the variances.

A modified model of (5) introduces the matrix V , which represents coherent noise sources. In this case $\vec{\varepsilon}_i$ contains only signals from incoherent noise guaranteeing that Σ_N can be expressed in form of (7).

$$\vec{X}_i = \underline{U}\vec{\beta}_i + \underline{V}\vec{\gamma}_i + \vec{\varepsilon}_i = [\underline{U} \quad \underline{V}] \begin{bmatrix} \vec{\beta}_i \\ \vec{\gamma}_i \end{bmatrix} + \vec{\varepsilon}_i = \underline{W}\vec{\alpha}_i + \vec{\varepsilon}_i \quad (8)$$

Then the eigenvectors of all dominant eigenvalues build the matrix W , which can be an arbitrary mixture of signal matrix U and coherent noise matrix V . In absence of coherent noise, two dominant eigenvalues exist representing the two natural source polarizations. In this case, the matrix W is equal to the matrix U and impedance tensor for each station can be derived. Existence of more than two dominant eigenvalues indicates that the MT data contain coherent noise and successful estimation of impedance tensor is only possible by separating coherent noise from the signal, which is not possible in every case.

With this knowledge we can implement a non-robust multiple-station approach in two steps. In the first step we calculate the incoherent noise variances for each channel to estimate the matrix Σ_N , while in the second step the eigenvalue problem is solved. The eigenvectors associated with the two largest eigenvalues build the matrix U . The impedance tensor Z_j for each station is derived by elements of U corresponding to the x and y components of electric and magnetic fields observed at station j .

$$Z_j = \begin{bmatrix} \zeta_{xj1} & \zeta_{xj2} \\ \zeta_{yj1} & \zeta_{yj2} \end{bmatrix} \begin{bmatrix} \eta_{xj1} & \eta_{xj2} \\ \eta_{yj1} & \eta_{yj2} \end{bmatrix}^{-1} \quad (9)$$

Application to MT data from Namibia

For the station network 103-109-702 two dominant eigenvalues exist almost over the entire frequency range (Figure 4). Therefore the two eigenvectors of these two eigenvalues are used to build the matrix U and the impedance tensor for each station. In Figure 5, robust single site results of station 109 are compared with results of the non-robust multiple-station approach. For most frequencies the results match very well. As expected the bias for the Z_{xx} component in the single site result is eliminated by the non-robust approach. The implemented approach works properly and in case of almost noise-free data the results are as good as or better than the robust single site results. Contain the data more noise than the non-robust approach does not suffice to resolve the impedance tensor over the entire frequency range. For some frequency of the second network, 100-102-110, more than two dominant eigenvalues exist. Using the eigenvectors associated with the two largest eigenvalues for estimation of impedance in such a case does not guarantee good results, because the matrix W is an arbitrary mixture of coherent noise signal and MT signal and the two largest eigenvalues do not need to represent the MT signal.

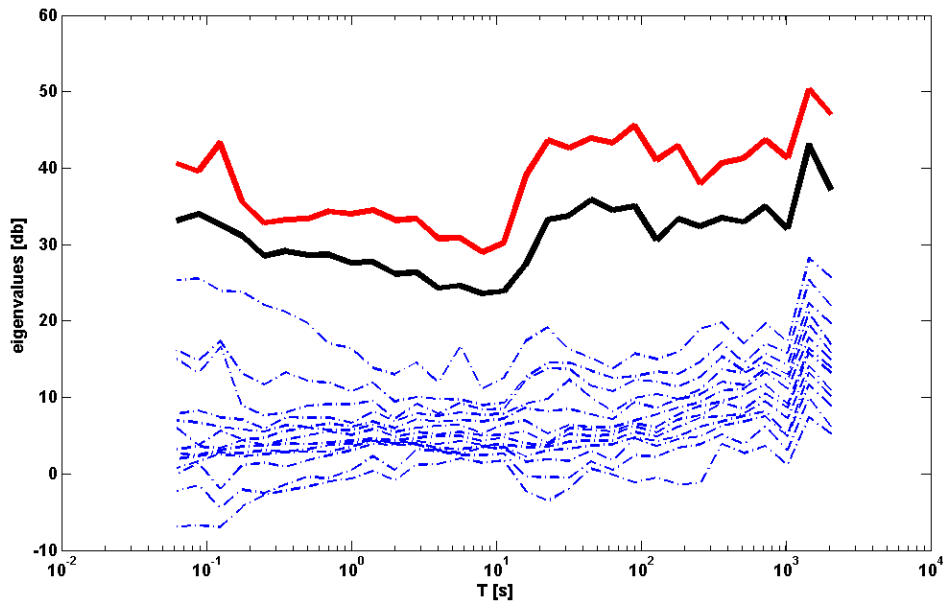


Figure 4: Eigenvalues for the non-robust multiple-station approach for station network 103-109-702. The two largest eigenvalues are marked in red and black and they are for almost all frequencies clearly larger than the other 13 eigenvalues. The eigenvectors associated with these two are used for estimation of impedance tensor. Magnitude of all eigenvalues is comparable with the eigenvalues shown by Egbert (1997).

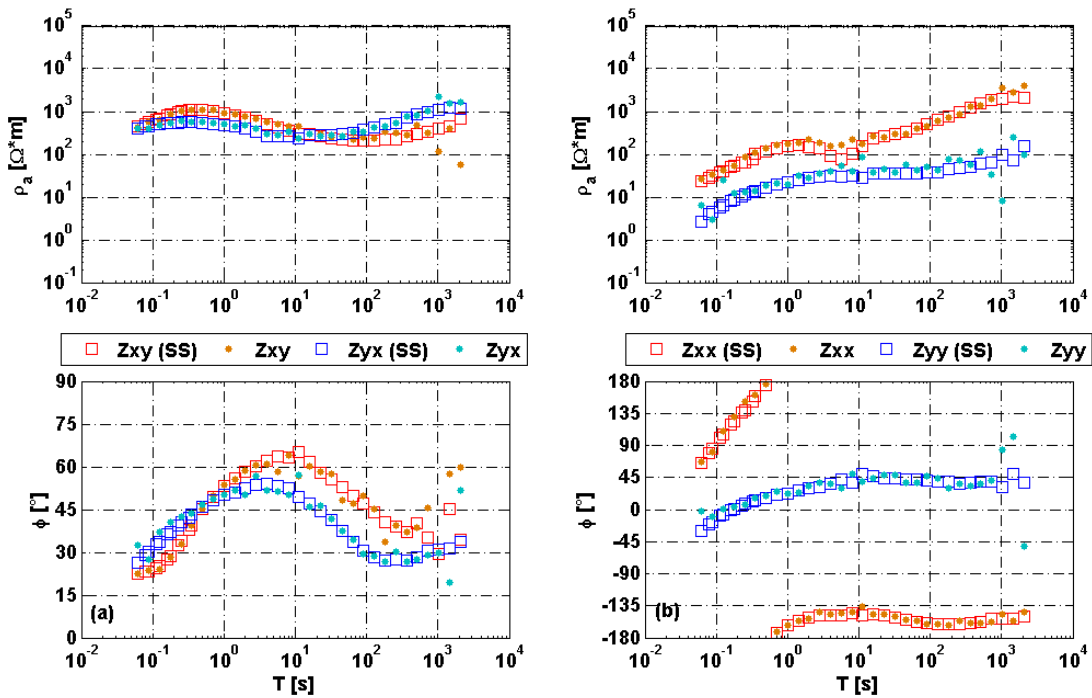


Figure 5: Comparison between robust single site results (red and blue squares) and non-robust multiple-station results (orange and cyan dots) for station 109. For most frequencies the non-robust multiple-station approach can resolve the curves in the same quality as robust single site processing. Only for long periods, where naturally less events exist and where we resolve more than two dominant eigenvalues, values scatter. The bias of component Z_{xx} is automatically eliminated by multiple-station approach (b).

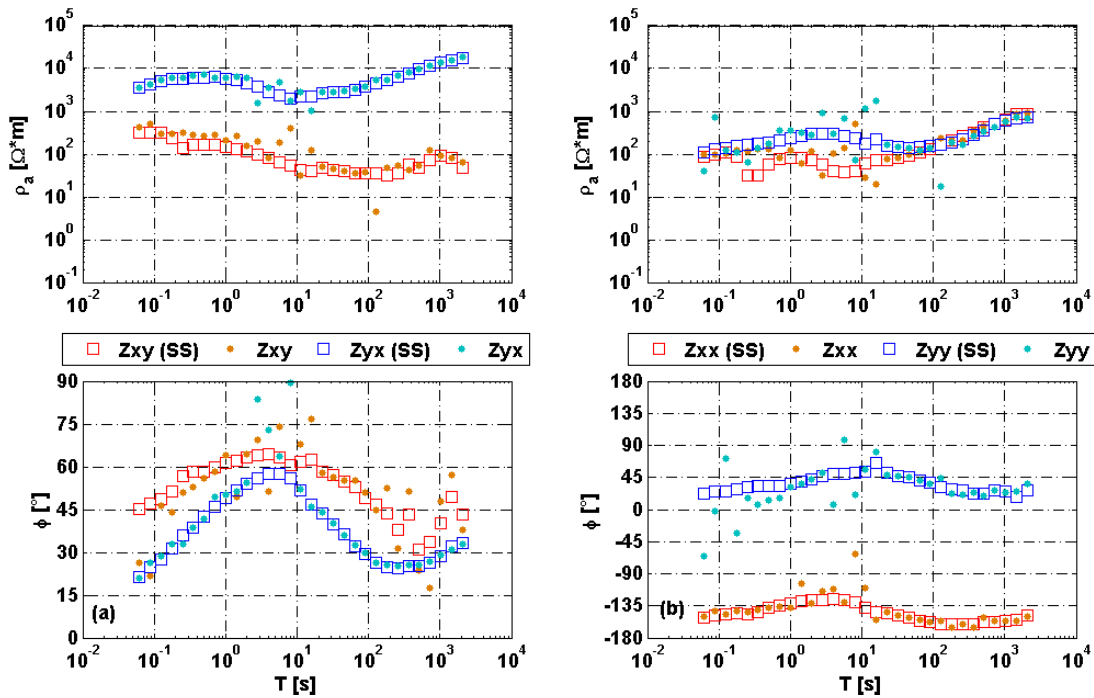


Figure 6: Comparison between robust single site results (red and blue squares) and non-robust multiple-station results (orange and cyan dots) for station 100. Single site results are obtained by using the robust algorithm with a threshold of 0.9 for the coherency criterion and without using phase criterion. For periods shorter than 10s more than two dominant eigenvalues exist and phase values for both off-diagonal (a) and diagonal components (b) scatter in this frequency range for the non-robust multiple-station approach. Although the general curve shapes are resolved for all four impedance tensor components, a robust algorithm is necessary to improve results of multiple-station approach.

Figure 6 shows an exemplary result for this network: Although for many frequencies the robust single site results and the non-robust multiple-station result match, a lot of frequencies exist with significant differences in the results. This particularly applies to the high frequencies for which four dominant eigenvalues are observed although data quality is good. Therefore the next step is an extension of the non-robust approach by robust statistics, which eliminate the influence of outliers.

COMBINATION WITH EXISTING ROBUST STATISTICS

The non-robust multiple-station approach is combined with parts of robust statistics which is already implemented in the processing scheme “EMERALD”. The robust algorithm consists of data selection criteria and a robust stacking algorithm described in more detail by e.g. Weckmann et al. (2005), Ritter et al. (1998) and Krings (2007). Normally it is used for standard single site and RR processing and therefore it is based on univariate and bivariate statistics e.g. by calculating bivariate coherences and using the bivariate equations in (1). Figure 7 shows results of the robust and non-robust multiple-station approach for station 109. Both results seem to coincide and illustrate that in this case the robust algorithm has no influence of the estimation of the impedance tensor. The main reason for this is the high data quality with almost no outliers within this station network.

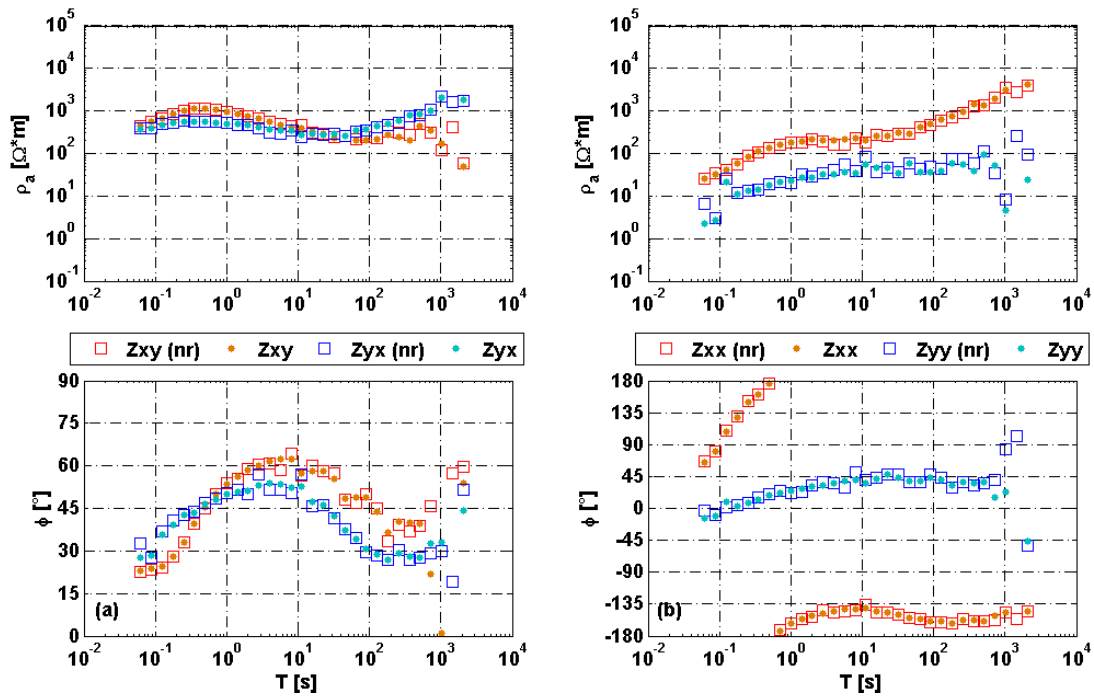


Figure 7: Comparison between non-robust multiple-station results (red and blue squares) and robust multiple-station results (orange and cyan dots) for station 109. For the robust processing coherency criterion with a threshold of 0.92 is used and the first part of the robust stacking algorithm using Huber and Tukey weights. Both results match well and in this case the robust algorithm seems to have hardly an influence of processing results.

The data of the second network contain more outliers, which normally should be sorted out by the robust algorithm. Unfortunately, this does not seem to be the case (Figure 8). The processing results of both approaches match well and a significant improvement of apparent resistivity and phase curves cannot be observed. The robust algorithm uses bivariate equations and statistics, which can not sufficiently characterize the multivariate structure of the dataset. Therefore a useful characterization of outliers is not possible by using bivariate statistics in combination with a multiple-station approach. Furthermore neither the non-robust nor the robust multiple-station approach separate coherent noise from desired MT signal. Processing of data, which contain coherent noise shown by a high number of dominant eigenvalues leads to unwanted results.

To illustrate advances of the multiple-station approach a comparison of non-robust single site and non-robust multiple-station processing of station 100 is shown in Figure 9. Single site processing without any robust algorithm and data selection criteria cannot resolve apparent resistivity and phase curves over the entire frequency range. On the other hand non-robust multiple-station processing can resolve the shape of the curves and the quality of the results is much higher.

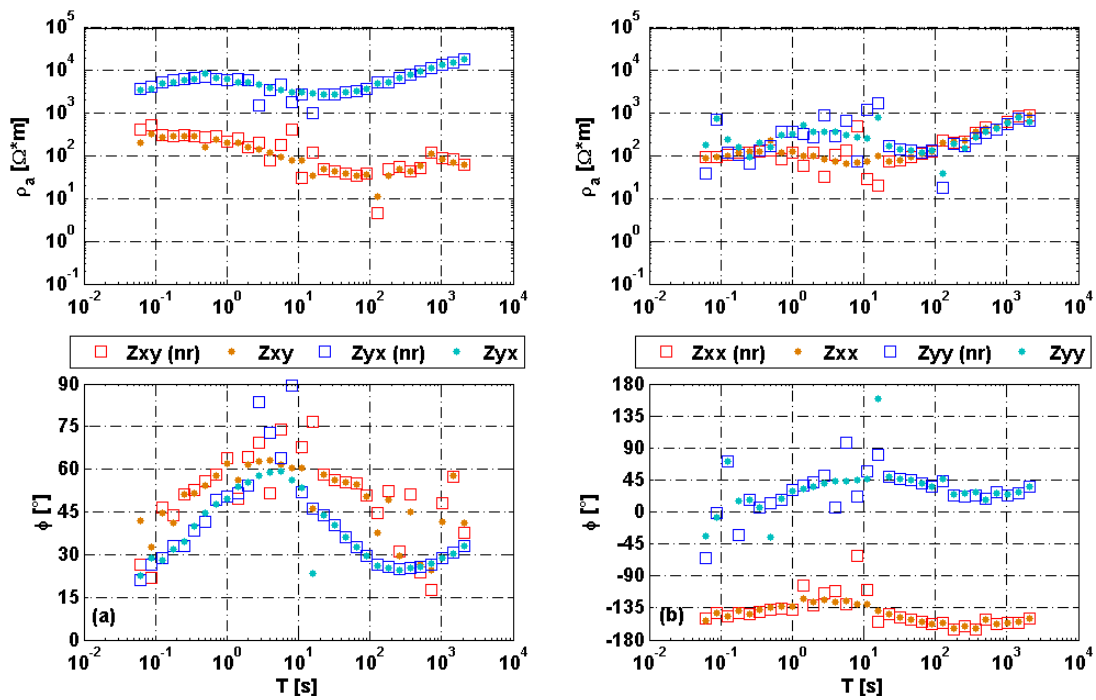


Figure 8: Comparison between non-robust multiple-station results (red and blue squares) and robust multiple-station results (orange and cyan dots) for station 100. For the robust processing coherency criterion with a threshold of 0.9 is used and the first part of the robust stacking algorithm using Huber and Tukey weights. Both results match for many frequencies. A significant improvement of apparent resistivity and phase curves cannot be observed.

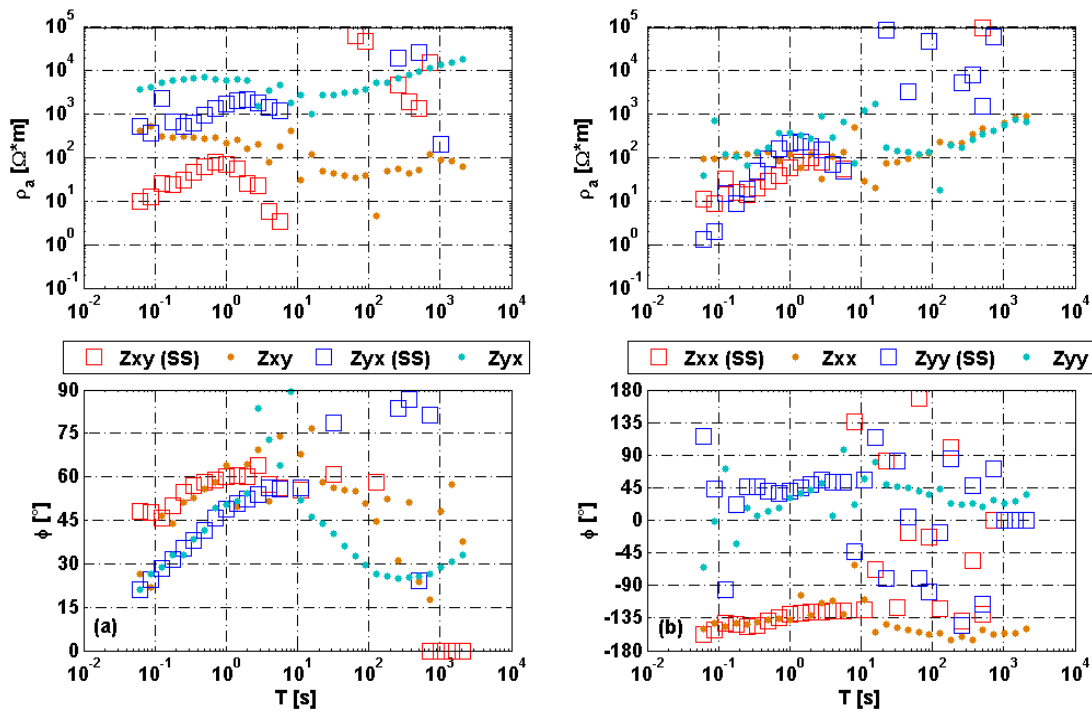


Figure 9: Comparison between non-robust single site results (red and blue squares) and non-robust multiple-station results (orange and cyan dots) for station 100. The non-robust multiple-station processing resolves the results much better than non-robust single site processing. In the period range between 10 – 100s some apparent resistivity values for the off-diagonal components (a) are higher than 10,0000 Ωm . These values are in this case unrealistic and are not displayed.

CONCLUSION

A non-robust multiple-station approach is integrated in our processing routines showing already improved results for almost noise-free data compared to robust single site or even remote reference estimators. However, in case of a higher level of man-made noise advances of the non-robust multiple-station approach in comparison with robust single-site results are not observed. But a direct comparison of non-robust single site and non-robust multiple-station processing shows the advances of the novel algorithm. Tests with the existing robust routines, which are based on bivariate assumptions, do not reveal a significant improvement. It seems that the robust algorithm based on bivariate assumptions counteracts the multivariate characteristic of the multiple-station approach. Therefore future steps include development of new data selection criteria and integration of robust multivariate statistics.

REFERENCES

Egbert G (1997) Robust multiple-station magnetotelluric data processing. *Geophys. J. Int.* 130, 475-496
 Krings T (2007) The influence of robust Statistics, Remote Reference, and Horizontal Magnetic Transfer Functions on data processing in Magnetotellurics. Diploma thesis, Westfälische Wilhelms-Universität, Münster

Ritter *et al.* (1998) New equipment and processing for magnetotelluric remote reference observations. *Geophysical Journal International* 132, 535-548

Weckmann *et al.* (2005) Effective noise separation for magnetotelluric single site data processing using a frequency domain selection scheme. *Geophysical Journal International* 161, 635-652